Integrating and sharing accession-level and omics-size genotype,

phenotype and environmental data: Experiences at the International Potato Center (CIP)

Simon, Reinhard
Cordova, Raúl A.
Hualla, Vilma R.
Salas, Elisa
Tejada, Sofia T.
Flores, Mirella R.
Pollreich, Severin
De Haan, Stef
Bonierbale, Merideth

International Potato Center,
CIP Headquarters
1558, Lima 12, Peru

r.simon@cgiar.org

Introduction

Plant breeding consists in the creation and selection of new genotypes. This involves not only keeping records across generations and environments but also accommodating data of increasing resolution on genotypes, phenotypes, and growth environments. Some such high-resolution characterization methods are Near-Infrared spectroscopy, metabolomics, next-generation sequencing and high resolution spatial-temporal-spectral photos. A first need is the integration and retrieval of this information. Such an integrated and complete set can be described in breeder's terms in six dimensions: a plant phenotype (P) is the result of a genotypes (G) interaction with its environment (E) given certain field management (M) practices. In addition, data on the administrative (A) context should be kept including staff involved, objectives and, if applicable, projects and donors; as well as on data documentation standards (S) like ontologies. The latter play an important part in exchanging and aggregating information. Here we describe the adoption of the 'Biomart' database for this purpose. While Biomart was developed originally to accommodate gene and sequencing data at a genomic scale we describe here how it can be used for breeding program data. This is being illustrated by current data warehousing in the potato breeding program at the International Potato Center (CIP). Particularly, genotype and phenotype can be transparently combined for further analysis in the decision process for the selection of new genotypes.

Materials and method

BioMart provides a free software and data services to the international scientific community. Following the BioMart conventions one DataMart was organized for potato, these included passport, characterization, evaluation (healthy tuber yield, late blight exposure trial, physiological and molecular), distribution data, soil and climate data (see Figure 1).
Best practices were followed to organize and design the web user interface: e.g. the visual interface BioMart was structured using the potato crop ontology (http://www.cropontology.org). The potato ontology is currently under development and is also used for data documentation. We also use a biomart feature to provide users with differential authorized access (see Figure 2).
Link to BioMart: http://germplasmdb.cip.cgiar.org/index.jsp

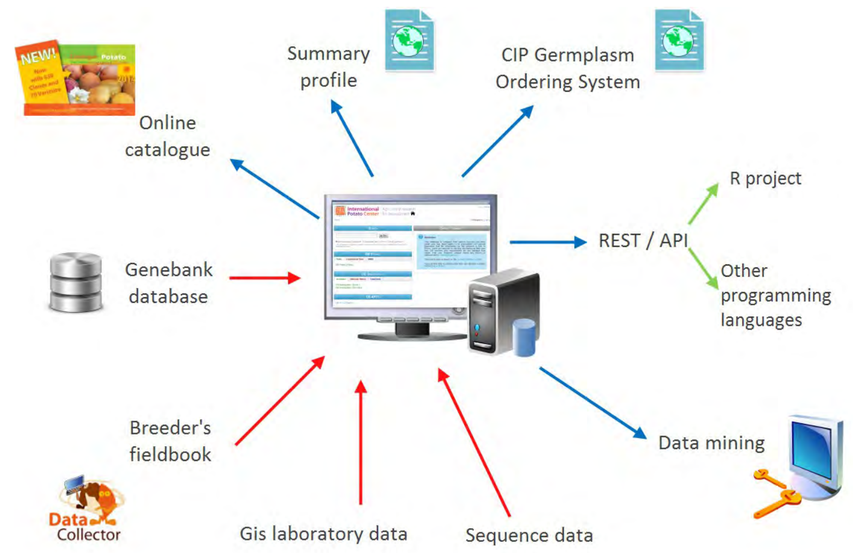

Figure 2: Research Tools Integration

Fig.1: Database structure and BioMart

Molecular Data

The database has 52 SSR primer pairs that cover the 12 potato chromosomes, the information is attached using the MIQAS format which considers information of population, pedigree, kinship, traits, markers, etc. Currently we have registered information on 22 primers for 80 advanced clones and 52 primers for 1564 genotype of potato germoplasm, both have 431,840 records in total.
Genomic and phenotypic information can be used for many purposes e.g analisys of the genome wide selection, genetic map,s etc. (see Figure 3).

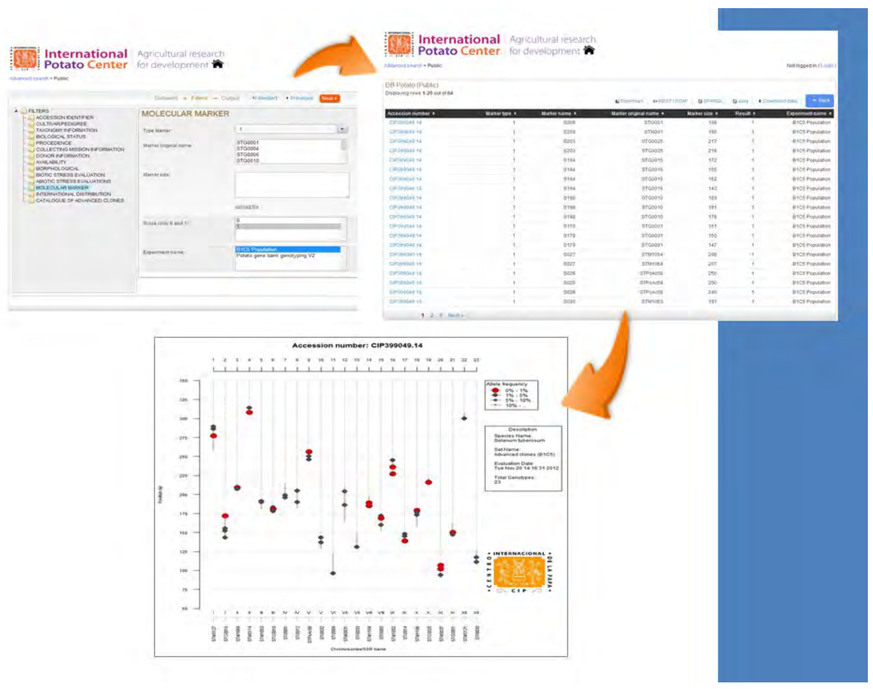

Figure 3: Graphic representation of molecular markers in BioMart

Results and Conclusions

"BioMart" has been used to store genomic, genetic, phenotype, pedigree, geographic , environmental data and metadata of the experiments (of breeders). It has been updated to currently 321 experiments from 9 different countries and 62 localities: biochemical evaluation (73 books), physiology (4 books) and healthy tuber yield (129 books). In addition, BioMart has historical information of: passport (47531 records), morphological trials (60419 records), international distribution (93054 records), daily weather (23 localities from 1974 to 2012 - 41,036 records), hourly weather (8 localities from 2004 to 2012 - 422,735 records) and soil (127 localities from 1989 to 2012 - 1,753 records).

Acknowledgment

CIP breeders Amoros, W.; Gastelo, M. , Mihovilovich, . For funding, the crop ontology group and the "Generation Challenge Program".

References
[1] www.biomart.org. Accessed on March, 19th 2014